Finding our foundation: Analysis of the LISA database for research retrievability

Carol Perryman MSLIS and Dihui Lu MS, MSIS

Abstract

Objective: The primary objective of this study is to examine the Library and Information Science Abstracts (LISA) database to determine if research literature can consistently be retrieved by using keywords identifying the research methodologies used.
Question: For the journals named, are articles identified as 'research' able to be consistently retrieved by using keywords related to research methodology?
Methods: Citations from the top 10 Library and Information Science journals for 2001 as identified by Koufogiannakis, Slater & Crumley (2004) were obtained, then a filter developed by Catherine Beverley (2004) was used independently to identify research articles. The resulting sets of citations were compared, then the two datasets were analyzed in order to consider retrievability and fit for assigned keywords from the LISA database.
Results: Although it would need to be tested against a random set of citations rather than the purposive sample tested here, our analysis suggests that retrieval using the descriptor terms alone may succeed in only 31.5% ± 5.2% of attempts, with a 95% confidence interval.
Conclusions: The LISA thesaurus is not consistent or sufficiently comprehensive to serve the needs of researchers. Recommendations for the improved retrievability of LIS research literature from the database are made.

Introduction

One of the problems mentioned by those who are leading the way in the Evidence-based Library and Information Practice (EBLIP) initiative is that of finding our own research in order to use it as a basis for decision-making. However, librarians do not need to be familiar with evidence-based concepts in order to be aware of the problems associated with locating relevant, current, and well-constructed decision support literature from our own resources. A quick look at library listservs will demonstrate that we tend to function on a peer-to-peer level, often conducting casual surveys to assemble a handful of experiential input from colleagues. It may be only when we attempt to gather more rigorous research-based information (in support of a grant, perhaps, or in assembling data to convince our stakeholders of the need for support) that we become aware that quite often, the data is not there – or if it is, that it is extremely difficult to retrieve from our own resources, such as the LISA database.

In justifying our study, we ask a 'catch-22' type of question: If there is little literature to support decision-making, why even look? If what is there cannot be found, how can it be used to expand our knowledge base, or to support our decisions in a practice environment? These questions are not focused upon the quality of research literature, which is not the immediate concern of this study. Instead, the primary objective of this work is to examine the LISA database in order to find out if the literature that is there can

consistently be retrieved by using descriptors that identify the research methodologies used. Secondly, the remainder of the keyword set (in the case of LISA, words from the title and abstract fields) will be scrutinized to see if there are dependable ways in which the searcher can identify appropriate literature. In doing so, we are building on the work of Koufogiannakis, Slater, and Crumley (Koufogiannakis, et al. 2004), who used manual review methods to compile a set of research articles, in order to identify the top journals for library and information science (LIS). Because the journals they identified are all indexed in the LISA database, we have used the citations identified by their study as the basic material for our research. A second dataset was extracted by use of a filter, or a pre-set search strategy, created by Catherine Beverley (2004) to identify research articles from the same database. We then compared retrieval between the two sets in order to analyze database performance, and also to evaluate and even build on the Beverley filter.

An additional, longer-term objective is that the authors hope to increase awareness of the problems associated with LIS resource access, and by doing so, to help bring about improvements. In the Medline database, the driving need to identify rigorous and valid studies for patient care as the result of Evidence-based Medicine became the impetus behind real improvements. Librarians and information professionals do not share an identical urgency, but they do have a need to move forward, and are impeded by barriers that can be resolved. Several of the assumptions providing the rationale for this study are derived from those which support Evidence-based Library and Information Practice initiatives concerning the need for increased rigor in research, as described by Booth and Brice (2004) and others.

First, outside of academic librarianship, research literature is not often used by practitioners because librarians tend to prefer peer-to-peer dissemination of information concerning issues in the LIS field, such as listservs and conferences in preference to research literature. Second, if research literature could be more easily identified by use of the existing tools, it might be used more often. Combined with the knowledge that LISA is the database most comprehensively indexing LIS journals, our rationale for using the CSA version of LISA was the unconfirmed conviction that for many librarians, DIALOG is a tool that is not widely available due, to cost. To be sure, the CSA LISA database may be nearly equally inaccessible, but the cost difference may mean it is more likely to be part of library subscriptions than DIALOG.

Background

While it is not our primary intent to debate the worth of conducting research in a practice setting, we cannot ignore the problems associated with doing so, because of the assumption stated, above, that our neglected resources are both a cause and effect of lack of use and usability. In a recent study of the use of library literature by LIS practitioners (Powell, Baker & Mika, 2002), members of ALA, ASIS, MLA and SLA who answered ‘no’ to a question of whether they read research articles were asked to consider why they did not do so. Participants answered that research-based literature did not seem relevant to their job; that the preference was for essay or opinion pieces; that they did not feel sufficiently knowledgeable in research methods, and that there was simply not enough

time. The study also found that while many practitioners conducted research, few actually published their findings, and confirmed similar barriers identified in earlier publications (Burdick, Doms, Dotie & Kinzie, 1990; Dalrymple & Fenske, 1992; Weaver, 1985). The benefits of conducting LIS research, and in particular, of using the LIS knowledge base, have been explored many times. Calling librarians the ideal proponents of innovation – the "catalyst(s) for new knowledge," Oliver (2000) argues for the involvement of community college librarians in action-based research. Zweizig (1999) supports the argument, adding that librarians are immersed in research in any number of ways: not only do we conduct research, we read studies done by others, and of course, we collect and provide access to the research output of all other disciplines. Williams and Winston (2003) examined the link between leadership competencies and the use of research and statistical methods in decision-making, and discussed the value and application of LIS research literature. Although their findings are particularly focused upon the academic library, the concerns are likely to be echoed in other settings. Among the desirable competencies, they argue, is the ability to find and use relevant data, because of the need for articulating a library vision that supports the college or university vision and that is consistent with the role of libraries in society requires library leaders at all levels to make informed decisions that contribute to organizational success and position the library for the future (Williams & Winston, 2003).

Addressing the professional development activities of beginning reference librarians, White mentions reading current research as an important way to generate ideas for further research (2001), while Poole (2000) adds that reading the literature, and adding to it by applying individual expertise to the base of research data on current practices can serve to make programs and services more effective. Discussing basic research methods, Watson-Boone (2000) pointsout that lacking information about what has come before, the librarian/researcher creates the possibility of wasting effort, introducing bias, and otherwise erring in ways that might have been prevented.

The best resources?

However, there are real concerns with regard to finding published 'best evidence' within our field, both because it is so sparse, and because a number of barriers exist to its retrieval. Addressing the issue of journal coverage in LIS databases, Jonathan Eldredge claims that

> "the library and informatics literature poses several unexpected challenges for the searcher. […] Coverage of any journal by one of these databases might suggest that the database with the most complete coverage represents the better choice. Yet, discrepancies across different years […] plus inconsistent coverage of any one journal within a single databases makes it difficult to recommend any one database to the busy practitioner" (Eldredge, 2004).

Problems associated with examining the published output of LIS professionals extend beyond the lack of awareness, support, skills in research, and the presence of practice

environments that do not always value, reward, or act on research. Access and retrieval barriers due to inconsistent indexing are recognized as transcending any single resource due to the lack of disciplinary boundaries, especially the practice of obtaining literature from many other areas of study. Those who have compared the coverage of major LIS resources remark that for the three largest and most commonly used (LISA, the Education Resources Information Center database (ERIC), and Library Literature), each has a separate vocabulary, and that there appear to be few overlaps in thesauri between them – unlike the more precise structure of Medical Subject Headings (MeSH) and Cinahl subject headings. In fact, note the authors, the lack of a common vocabulary for indexing is one "inherent in the nature of librarianship" (Stieg & Atkinson, 1988). Identifying the top journals in LIS appears to also be problematic, due to lack of agreement among researchers on methods for doing so.

Williams and Winston (2003) conducted a study of citation patterns in LIS research, selecting the top five journals identified by the Institute for Scientific Information (ISI) journal citation reports for 2002 (College & Research Libraries, Library Quarterly, Library Resources and Technical Services, Library Trends, and Library & Information Science Research). To do so, they discarded information science titles and focused solely on library science. The choice of ISI as a means by which top LIS titles might be identified seems questionable given the number of LIS journals not indexed in that resource: 15 of the top 30 journals identified in the study by Koufogiannakis, et al.(2004) are missing from their coverage.

In the absence of any existing citation index such as ISI to track publishing trends, the existing LIS databases, themselves, must serve as raw data for analysis. Identifying the top journals for research (in terms of frequency) was the partial purpose of Koufogiannakis, et al.(2004), who tested the domains for LIS research conceptualized by an earlier study by Crumley and Koufogiannakis (2002). To do so, Koufogiannakis, et al. reviewed 217 refereed LIS journals, determining which articles for the 2001 publications of these journals met the criteria of being 'research,' defined as "an inquiry, which is carried out, at least to some degree, by a systematic method with the purpose of eliciting some new facts, concepts or ideas" (Peritz, 1980). We note here that the study deliberately excluded journals whose content fell more to the information science area than the library one. Of the 2,664 articles examined, 807 (30.3%) met the criteria for classification as research. Ranking journals by volume of research articles published in 2001 allowed the authors to derive their 'top ten' LIS journals. It is this set of articles which we used as our own basic data.

The LISA database and its indexing

A study comparing the Dialog versions of Information Science Abstracts (ISA), LISA, and Library Literature (Read & Smith, 2000) considered coverage and overlap in subject areas of concern to LIS professionals. Their conclusion was that Library Literature led the three in terms of overall retrieval, with ISA running third, but that of the three, LISA had the largest backfile (from 1969) and that furthermore, overlap between two top

databases was never more than 21%. For comprehensive coverage, they concluded, librarians must expect to search more than one resource.

A number of studies have investigated the quality and structure of LISA, mostly by comparing its coverage and indexing consistency against other LIS databases, although we found none specifically addressing the CSA version. Jacsó mentions that he finds the CSA version to be a vast improvement over others (1997), but his focus for analysis is on comparisons between the Dialog versions of LISA, Library Literature, and ISA. In a chapter discussing quality of subject indexing (2001), he mentions that the transition from print to computer-based resources, subject term accuracy appears to have lost some urgency, due to the availability of keyword and even full-text searchability. Jacsó further points out that for purposes of conducting a comparison between DIALOG's ISA and the LISA database, the title field proves to be a more consistently fruitful search due to inconsistently applied vocabulary in the abstract and descriptor fields. He refers to the creation of a 'real' thesaurus post-purchase by Bowker-Saur in 1992, but mentions that though this helped to correct the data quality problems he had identified, the improvements were not retrospective, making the descriptor field unreliable. In a 1992 study, Jacsó describes the extreme redundancy of descriptor terms used in LISA, expressing the concern that it has the effect of reducing precision in retrieval, and that such occurrences are a strong indicator of the use of machine-generated methods to build the index, because it would be unlikely that humans would make such errors (1992).

Hood and Wilson (1994) used as their data "all of the 11453 records of the 1969 to September 1991 edition of LISA on CD-Rom" to examine the database indexing practices, describing the 'chaining' method employed at that time. Their findings included numerous errors in the date added (DA), accession number (AN), and descriptor (DE) fields. Most of the discrepancies described appear to have been corrected by the time we queried the CSA LISA database for the examples used in their article (an exception being that there are articles indexed under the 'USA' descriptor not retrieval by use of the standard 'U.S.A.' descriptor) (Hood & Wilson, 1994). The authors furthermore refer to a statement by LISA's editor, Moore (1988) that 'the removal of all systems and organizations from the subject index (DE) would be effective from January 1979 onwards," a statement that we can contradict with some confidence.

Other problems recognized by researchers include discrepancies in spelling, such as the use of both singular and plural terms (Jacsó, 1997), and format and case of terms (Hood & Wilson, 1994). A survey of authors done by Stieg and Atkinson (1988) asking for assessment of the descriptors assigned by the LISA indexers found that authors objected to lack of specificity, the use of too-general terms, lack of accuracy, lack of comprehensiveness (e.g., important concepts were not included in the descriptor groups, and there was a lack of descriptors for concepts concerning the orientation or approach to the research study).

Methods

Article citations that form the core part of this inquiry were provided by Koufogiannakis, Slater and Crumley, who generously sent their entire Procite database (personal correspondence, D. Koufogiannakis, L. Slater, & E. Crumley, May 25, 2005). The dataset was selected for our purposes because it comprised a manually selected set of research literature from the library profession, with clearly delineated boundaries of year and availability through the LISA database. Manual selection has long been the 'gold standard' for retrieval, so the records provide a natural starting point for analysis.

A second set of citations was retrieved by use of a 'research' filter adapted from one built by Beverley (2004) to restrict retrieval to the top 10 journals identified by Koufogiannakis, et al, and to a publication date of 2001, as they had also done. The filter is the result of years of trial and error testing done by Catherine Beverley and colleagues at the School of Health and Related Research (ScHARR) and the Department of Information Studies in Sheffield, England (C. Beverley (personal communication, October 25, 2005). The first part of our search statement () is the combined set of top LIS journal titles (Koufogiannakis, et al. 2004), while the underlined text at the end of the query is the filter built by Beverley (2004) to retrieve research articles from the SilverPlatter LISA database, adapted slightly by the authors for use with the CSA version of LISA. Problems were encountered in entering the entire filter as one search string. The authors received messages indicating zero retrieval, so the search string had to be divided, then recombined for a total result.

((JN=journal of the american society for information science) or (JN=scientometrics) or (JN=information processing and management) or (JN=information processing & management) or (JN=(((college & research libraries) or JN=(college and research libraries)) NOT (college & research libraries news))) or (JN=(journal of library administration)) or (JN=(bulletin of the medical library association)) or (JN=(libraries & culture)) or (JN=(journal of documentation)) or (JN=(journal of information science)) or (JN=(journal of academic librarianship))) and ((DE=(research or survey* or evaluation)) or (TI=(research* or methodolog* or hypothes* or experiment* or comparison* or comparative* or interview* or survey* or questionnaire* or focus group or qualitative* or quantitative* or grounded theory or ethnograph* or data analy*)) or (AB=(research* or methodolog* or hypothes* or experiment* or comparison* or comparative* or interview* or survey* or questionnaire* or focus group or qualitative* or quantitative* or grounded theory or ethnograph* or data analy*)) Date Range: 2001 to 2001

Table 1. The adapted Beverley 'research' filter (underlined text)(Beverley, 2004) combined with journal titles and year limits used in the Koufogiannakis study (Koufogiannakis, et al. 2004).

Using the "research" filter built by Beverley, we retrieved 365 articles from the LISA database for the top 10 journals. We also obtained 311 citations identified as research articles reviewed by Koufogiannakis, et al. using manual methods. The sets of citations (comprised of the LISA identifying number (AN), author (AU), title (TI), journal (JN), publication date (PD), keywords (KW), descriptors (DE), abstract (AB), volume (JV), issue (JI), and page numbers (JP) were populated into a MySQL database, then tagged to

ensure that the records would continue to be identified as the results of the Beverley filter retrieval or the Koufogiannakis manual retrieval set. A Perl script was used to automatically parse the citation file from the Beverley filter and from the ProCite data provided by Koufogiannakis et al. MySQL (a relational database) was used instead of a flat file such as Microsoft Excel to better assure data integrity and consistency. Figure 1 shows the simple schema we used to store and analyze our citations.

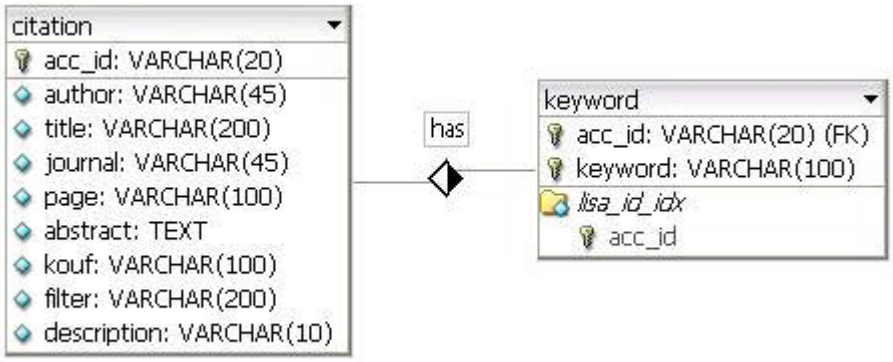


Figure 1: A simple schema to store citations and their descriptors.

Two tables are used in our database: citation and keyword. The citation table stores bibliographic information for each unique citation while the keyword table stores descriptors for each citation. Each citation has a unique LISA accession number (acc_id) and other attributes like author, title, journal name, paging information and abstract etc. One citation has multiple descriptors. The unique LISA accession number together with keyword decides the uniqueness of each descriptor since many citations could share the same descriptor. Table 2 below shows the number of citations we populated into the MySql database.

Results

After reexamining the dataset from Koufogiannakis, Slater and Crumley (Koufogiannakis, et al. 2004), we identified 305 total unique citations. 206 citations are identified by both Koufogiannakis, et al. AND the Beverley filter (Kouf∪Filter), while a total of 464 citations were retrieved by combining the two data sets (Kouf∩Filter).

Journal name	K	B	K∪B	K∩B
Bulletin of the Medical Library Association	23	23	29	17
College and Research Libraries	28	30	32	26
Information Processing and Management	31	26	37	20
Journal of Documentation	18	22	26	14
Journal of Information Science	17	17	20	14
Journal of Library Administration	25	97	98	24
Journal of the American Society for Information	68	53	82	39
Libraries and Culture	23	3	25	1
Scientometrics	55	66	85	36
The Journal of Academic Librarianship	17	28	30	15
Total	**305**	**365**	**464**	**206**

Table 2. Number of citations for the top 10 journals in the LISA database.
Kouf ∪Filter: the total number of citations from either by Beverley filter retrieval OR Koufogiannakis dataset.
Kouf: The number of citations from the Koufogiannakis dataset alone.
Filter: The number of citations retrieved by the Beverley filter alone.
Kouf∩Filter: The total number of citations retrieved by the Beverly filter AND identified by Koufogiannakis.

The Koufogiannakis set, representing the 'gold standard' of manual retrieval, was used as the baseline for comparison. The 305 research articles identified by Koufogiannakis, Slater & Crumley's study (Koufogiannakis, et al. 2004) were compared with articles retrieved by use of the amended Beverley filter (2004) to test if a filter built with research related keywords/descriptors could re-retrieve the same research literature that was identified by human experts. We did two runs of filter retrievals. First, we constructed a search ANDing the Beverley filter descriptor terms (DE=(research or survey* or evaluation) against the baseline set to examine the Beverley-assigned descriptor term occurrence. 98 of 311 (31.1%) of the Koufogiannakis-identified research citations were retrieved using these two descriptor terms. Although it would need to be tested against a random set of citations rather than the purposive sample population tested here, our analysis suggests that retrieval using the descriptor terms alone may succeed in 31.5% ± 5.2% of attempts, with a 95% confidence interval.

Second, expanding to the entire keyword set (title, abstract, and descriptors) making up the Beverley filter, we find that 206 of 311 (66%) of the baseline (Koufogiannakis) set was retrieved. Taking into account that the baseline set is not random, we find that research articles may be retrieved on a consistent basis in 66.2% ± 5.6% of attempts using this filter, with a 95% confidence interval.

Search	Total citations	Total number of descriptors	Total number of unique descriptors	K='Y' and B='Y'	K='Y' and B='N'	K='N' and B='Y'
1. Koufogiannakis (K) set combined with all Beverly (B) filter terms (B1) (Descriptor+title+abstract)	464	2988	798	206	99	159
2. Koufogiannakis (K) set combined with only Beverly descriptor terms (B2)				98	211	84

Table 3. Comparison of Koufogiannakis citations with results from two runs of Beverley filter.

The first Beverley filter was built with research related keywords querying against descriptors, titles and abstracts in LISA and the second filter queries were built with research related descriptors only. Table 3 shows that a filtered query against descriptor, title and abstract fields retrieved 68% (206/305) of the research literature identified by human experts, while one using research related descriptors alone retrieved only 32% (98/305) of the research articles. This result further confirms that the LISA thesaurus is not sufficient for researchers to retrieve research literature, and that using research-related keywords provides more comprehensive search results. In addition, for the 464 unique citations from both the Koufogiannakis dataset and the Beverley filter, there are a total of 2988 descriptors and 798 unique descriptors. On average, each citation has 6.4 descriptors (2988/464).

Next, a descriptor-frequency ranking table was generated by querying our database (table 4):

B	Frequency	K	Frequency
Research	100	Online information retrieval	54
Into	92	And	52
OCLC	58	Searching	49
World Wide Web	55	Of	48
Scientometrics	53	World Wide Web	46
And	52	Scientometrics	43
Internet	49	Research	43
USA	48	University libraries	37
Evaluation	45	Into	36

Online information retrieval	44	Internet	34
Searching	42	Periodicals	33
University libraries	38	USA	33
Of	37	User surveys	30
Students	32	Students	29
Academic libraries	31	Academic libraries	28
Electronic media	31	Citation analysis	27
User surveys	30	Evaluation	26
Periodicals	25	Libraries	23
Library management	23	History	22
Libraries	23	Users	20
Surveys	22	OCLC	20
Users	22	Articles	20
Research libraries	20	Surveys	19
Citation analysis	19	Bibliometrics	16
Medical libraries	18	Electronic media	16
Articles	18	Cold War period	15
Performance measures	17	Medical libraries	15
Use for	16	Web sites	14
Science	16	Science	14
Library staff	15	Information seeking behaviour	14
Medicine	15	Use for	14
Bibliometrics	14	Medicine	14
Web sites	14	Use	13
Library materials	14	Library staff	12
Use	13	Library materials	11
Electronic periodicals	12	Comparative studies	10
Formats	12	Search engines	9
Information seeking behaviour	12	Automatic text analysis	9
Collaboration	12	User training	9
Image databases	11	Publishing	9
Electronic publishing	11	Search strategies	9
Distance learning	10	UK	8
User training	10	Impact factors	8
Influence on	10	TREC	8
Search engines	9	User needs	8
Software	9	Influence on	8

Science and technology	9	Image databases	8
Comparative studies	9	Library management	7
User services	8	Faculty	7

Table 4. Descriptor-frequency ranking table for the combined dataset from the Beverley filter and Koufogiannakis sets, respectively (top 49 records of the total 2998).

The entire table of 2998 terms and phrases was scrutinized in order to identify those related to research methodology (shown in Table 5), so that conclusions could be drawn in answer to the initial research question, 'are articles identified as 'research' able to be consistently retrieved by using descriptors related to research methodology?' However, we note that without examining the articles themselves, there is no consistently reliable way to know whether the subject is, for example, how to conduct surveys, or whether surveys were used in conducting research.

Furthermore, based upon the lack of consistency with which they are found in any of the keyword fields, we suggest that these terms should be applied only with the understanding that their retrieval may be neither comprehensive nor specific. Interesting discrepancies include the variation between the use of the indicator term 'research' between the two sets of data, and the inclusion of both 'evaluation' and 'library evaluation.'

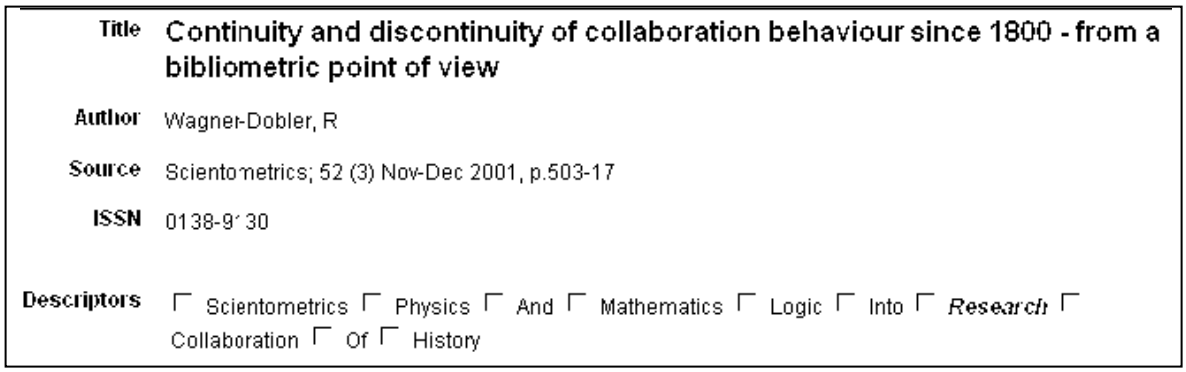
Descriptors	B	K
bibliometrics	14	16
citation analysis	19	27
cluster analysis	0	1
comparative studies	9	10
comparison with	4	4
content analysis	1	0
data mining	3	1
evaluation	45	26
exploratory data analysis	1	0
informetrics	3	3
interviews	1	1
library evaluation	1	0
linguistic analysis	1	3
quantitative analysis	1	1
quantitative methods	1	0
research	100	43
research methods	6	4
scientometrics	53	43
semantic analysis	1	1

surveys	22	19
systematic reviews	2	2
text analysis	2	2
user surveys	30	30
Total	**222**	**246**

Table 5. Research methodology-related descriptor terms identified from the ranked descriptor set of terms for the Koufogiannakis (K) and the Beverley filter (F) retrieval datasets. Numbers indicate the frequency with which the associated descriptor terms were used for the Koufogiannakis and Beverley sets, respectively.

In the case of the 'K' (Koufogiannakis) set, research-related descriptors were used for 80.6% of the records (246/305), and for the 'F' (Beverley filter) set, they were used for 60.8% of the records (222/365). By examining the entire ranked listing of keywords, we compiled a set of terms that could potentially improve retrieval even above the present Beverley filter set: analy*, assess*, conclud*, compar*, discover*, evaluat*, examin*, recommend*, stud*, and suggest*. We did not test these terms against the baseline set, because the terms were derived from it. These are words that we felt did not directly represent a research methodology, but that appeared to be indicators of research content. Such terms can be appended to a search in order to increase retrieval, but their use is likely to increase the number of 'false drops,' or inappropriate retrieval.

Our scrutiny of this descriptor-frequency ranking list also indicates that the LISA descriptor set appears to include terms generated by computer programmed frequency rankings, rather than being the result of human selection. Evidence of this can be seen in the above table (table 6), which shows a number of non-descriptive terms used as descriptors for records in those we examined. Among those we noted beyond this listing are the following terms and phrases: implications for, on, authors, role in, to, and, viewpoint on, theories, connected with, unobtrusive evaluation of reference service and individual responsibility, FEATURES, Fred, number, questions, portrayal in et al. Unfortunately, many of the assigned terms lack the specificity that is needed to retrieve pertinent content, as can be seen in Figure 2. The same record also includes four descriptor terms that are questionable at best, 'and,' 'logic,' 'into,' and 'of' – four of the ten descriptors provided that do nothing to describe content. Looking at the 'value added' ability to select from above the descriptor terms in order to build a search, one wonders how a search structured from such selections would be helpful in any way. In fact, a restructured search using the descriptor terms 'and,' 'into,' and 'of' netted the authors 17 conceptually discrepant records, which is hardly surprising.



Title	**Continuity and discontinuity of collaboration behaviour since 1800 - from a bibliometric point of view**
Author	Wagner-Dobler, R
Source	Scientometrics; 52 (3) Nov-Dec 2001, p.503-17
ISSN	0138-9130
Descriptors	Scientometrics Physics And Mathematics Logic Into *Research* Collaboration Of History

Figure 2. Example of one of the records obtained by searching for non-essential descriptor terms such as 'and' and 'into.' Note that although the title indicates this to be a bibliometric study – and in fact, that 'bibliometrics' is a term used in the thesaurus, it is not assigned as a descriptor.

'Sampling techniques' may be found, but not through 'data collection'; its listing provides neither narrower nor broader conceptual term links. This inconsistency is a real problem, since a descriptor search with that term (DE=sampling techniques) retrieved 56 articles whose topics include bibliometric studies, systems development using sampling techniques for analysis, an availability study for books in a library, and a way to conduct journal cost studies. Because terms like 'sampling techniques' are available, they can be used to retrieve records. However, the individual attempting to build a search using this thesaurus will find themselves bewildered by the logic, especially since it is not explained. There are no definitions provided for terms or phrases in the thesaurus, again leaving the searcher to guess at what might have been meant by terms such as 'aboutness.'

Other problems observed during analysis include those others have noted, such as misspellings (e.g.: the use of 'interpretative' instead of 'interpretive'). In direct contradiction of Moore's 1991 assertion that names of systems and organizations would be effective from 1979 onward (in Hood & Wilson), we point to the use of descriptors 'OCLC' and 'TREC' in the records we retrieved, all 2001 articles. Furthermore, and frustratingly, 'TREC' is nowhere to be seen in the LISA thesaurus, though the term 'OCLC' and its various permutations is evident. An entire issue of JASIS is missing from the LISA database (volume 52 issue 10, 2001).

The LISA thesaurus does not appear to be complete, sufficiently comprehensive, or consistently applied enough to serve the needs of researchers. Looking at the term 'research methods' itself within the hierarchical display, one finds a very limited set of descriptors. These terms do not include quantitative research, evaluation, experiment, comparison, and most of the terms one would expect to find under that broader heading, meaning that the searcher must also employ creative alternatives such as the Beverley

filter to retrieve literature. One of the narrower terms, 'data collection' includes only the descriptor 'transaction logs,' and not surveys, interviews, or other terms logically within that category. The term 'surveys' and its narrower terms may indeed be found within the descriptor terms, but though it links back to 'research,' it does not appear to be linked to, from 'research methods.'

Conclusions

"Does it matter?" is a question that should be asked of every research study. Does it matter that LISA is an imperfect tool, that access to our own foundation is impeded by indexing, and by our own use of non-standardized vocabularies? Why should librarians be concerned with LIS research literature, when our top-priority need is to serve our communities, and to become familiar with the vocabularies and literature of those populations? Our own answer is yes: LIS literature and research matter. But the question is one we would also direct to the library community, and to library communities, more generally. Change cannot occur unless there is agreement that a problem exists, and that its recognition and alleviation concern us all.

Our examination of the LISA database demonstrates that neither indexer-supplied descriptor terms, nor author-supplied keywords can reliably serve to identify research methodologies, at present. The implications for practice are several. First, the user is very likely to be frustrated in their attempt to gather supportive information for their own use. Second, the use of proven methods for searching, such as 'pearl gathering' (finding a pertinent article, then building a search from its index terms and keywords), will not work in this database, due to pervasive inconsistencies found in LISA, and in our own vocabularies. Third, although the user can improve on retrieval by using the Beverley filter (2004), and by judicious use of the terms we have suggested, the likelihood of obtaining false hits will increase in both cases.

Further studies looking at issues of accessibility for the tools of our own profession are highly recommended, and should be of direct concern to library associations committed to helping foster use and creation of a more robust body of research literature. Although looking at the plethora of earlier research examining the quality and consistency of LISA's indexing in its various incarnations indicates that there has been considerable improvement, there are still many serious issues to address if this database is to fulfill its goal as the primary repository for library literature. The lack of scope notes for descriptor terms does real disservice to the user, confounded by what seem to be unstandardized terms used in the descriptor fields, which further points to the glaring need for quality assurance procedures. Other research worth pursuing involves the quality assessment of research articles, once they are able to be identified, using the methods described by pioneers in the evidence-based information initiative (Booth & Brice, 2004), and the development of citation reports similar to those provided by ISI, due to that resource's lack of coverage.

Koufogiannakis, et al.(2004) have moved the profession forward by identifying domains of practice, which can only help with indexing and retrieval. Their identification of the

top research journals for LIS based upon the number of research articles published is also an excellent step toward strengthening our knowledge base. The filters built by Catherine Beverley (2004) add to our research 'toolbox,' as do the growing number of evidence-based studies being published. Crucial elements of professionalism are the existence of a solid research base and established links between published research and the practice environment. The work of Stieg & Atkinson (1988) concludes with a call for a national repository similar to the National Library of Medicine's MedLine, and though we echo that call, we would add to it by suggesting that it should be more global than national in scope. We would further wish for improved access to existing repositories, especially LISA, and suggest that a careful examination of the quality assurance processes used in indexing should occur. The inclusion by LISA indexers of domain-specific terms, such as those suggested by the work of Koufogiannakis, et al.(2004), and the involvement of editors for LIS journals in requiring structured abstracts and author-supplied keywords, would work toward ensuring the continued value of our largest repository of library knowledge.

Bibliography

1. Beverley, C. (2004). Searching the library and information science literature. In Booth, A., & Brice, A. (Eds.), Evidence-based practice for information professionals: A handbook (pp.89-103). London: Facet Publications.
2. Booth, A., & Brice, A. (Eds.). (2004). Evidence-based practice for information professionals: A handbook. London: Facet Publications, 2004.
3. Burdick, A.J., Doms, C.A., Doty, C.C., & Kinzie, L.A. (1990, October). Research activities among health sciences librarians: A survey. Bulletin of the Medical Library Associaton, 78, 400–402.
4. Crumley, E., & Koufogiannakis, D. (2002). Developing evidence-based librarianship: Practical steps for implementation. Health Information & Libraries Journal, 19, 61-70.
5. Dalrymple, P.W., & Fenske, R.E. (1992, October). Factors influencing research productivity among health sciences librarians. Bulletin of the Medical Library Association, 80, 353-360.
6. Eldredge, J. (2004). How good is the evidence base? In Booth, A., & Brice, A. (Eds.), Evidence-based practice for information professionals: A handbook (pp.36-37). London: Facet Publications.
7. Hood, W., & Wilson, C.S. (1994). Indexing terms in the LISA database on CD-Rom. Information Processing & Management, 30, 327-342.
8. Jacsó, P. (1992). CD-ROM Software, dataware, and hardware: Evaluation, selection, and installation. Englewood, CO: Libraries Unlimited.
9. Jacsó, P. (1997). Content evaluation of databases. Annual Review of Information Science and Technology, 32, 231-267.
10. Jacsó, P. (2001). Content evaluation of textual CD-Rom and web databases.Englewood, CO: Libraries Unlimited.
11. Koufogiannakis, D., Slater, L., & Crumley, E. (2004). A content analysis of librarianship research. Journal of Information Science, 30, 227-239.
12. Moore, N.L. (1988, April). LISA indexing: Economic aspects of controlled indexing.

Indexer, 16, 11-16.
13. Oliver, S. (2000, March). Revolutionizing how we generate new knowledge: A challenge for librarians, health professionals, service users and researchers. Health Libraries Review, 17, 22-5.
14. Peritz, B.C. (1980, Fall). The methods of library science research: Some results from a bibliometric survey. Library Research, 2, 251-68.
15. Poole, C.E. (2000, September). Importance of research and publication by community college librarians. College & Research Libraries, 61, 486-9.
16. Powell, R.R., Baker, L.M., & Mika, J.J. (2002). Library and information science practitioners and research. Library & Information Science Research, 24, 49-72.
17. Read, E.J., & Smith, R.C. (2000). Searching for library and information science literature: A comparison of coverage in three databases. Bulletin of the Medical Library Association, 19,118-27.
18. Stieg, M.F., & Atkinson, J.L. (1988, October). Librarianship online: Old problems, no new solutions. Library Journal, 113, 48-59.
19. Watson-Boone, R. (2000, March). Academic librarians as practitioner-researchers. Journal ofAcademic Librarianship, 26, 85-93.
20. Weaver, C.G. (1985, April). Writing for publication: Barriers and bridges. Bulletin of the Medical Library Association, 73, 213-4.
21. White, G. (2001). The professional development of reference librarians: Implications of research, publication, and service. Reference Librarian, 35, 337-350.
22. Williams, J.F., & Winston, M.D. (2003). Leadership competencies and the importance of research methods and statistical analysis in decision-making and research and publication: A study of citation patterns. Library & Information Science Research, 25, 387-402.
23. Zweizig, D.L. (1999). Support to librarians in their research roles. Library Quarterly, 69, 363-366.